TNO-report
TM-96-A051

title

Image fusion improves situational awareness

TNO Human Factors
Research Institute

Kampweg 5
P.O. Box 23
3769 ZG Soesterberg
The Netherlands

Phone +31 346 35 62 11
Fax +31 346 35 39 77

authors

A. Toet
J.K. IJspeert
M.J. van Dorresteijn

date

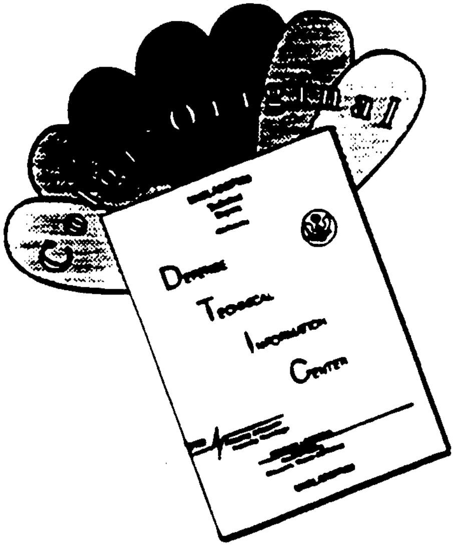
31 October 1996

number of pages : 28 (incl. appendices,
excl. distribution list)

TNO Human Factors Research Institute
is part of TNO Defence research
to which also belong:
TNO Physics and Electronics Laboratory
TNO Prins Maurits Laboratory

Netherlands Organization for
Applied Scientific Research

DISCLAIMER NOTICE



REPORT DOCUMENTATION PAGE

1. DEFENCE REPORT NUMBER (MOD-NL) RP 96-0192	2. RECIPIENT'S ACCESSION NUMBER	3. PERFORMING ORGANIZATION REPORT NUMBER TM-96-A051
4. PROJECT/TASK/WORK UNIT NO. 786.1	5. CONTRACT NUMBER A96/KL/347	6. REPORT DATE 31 October 1996
7. NUMBER OF PAGES 28	8. NUMBER OF REFERENCES 25	9. TYPE OF REPORT AND DATES COVERED Final

10. TITLE AND SUBTITLE

Image fusion improves situational awareness

11. AUTHOR(S)

A. Toet, J.K. IJspeert and M.J. van Dorresteijn

12. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

TNO Human Factors Research Institute
Kampweg 5
3769 DE SOESTERBERG

13. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Director of Army Research and Development
Van der Burchlaan 31
2597 PC DEN HAAG

14. SUPPLEMENTARY NOTES

15. ABSTRACT (MAXIMUM 200 WORDS, 1044 BYTE)

Two recently developed false colour image fusion techniques, the TNO fusion scheme (Toet & Walraven, 1996) and the MIT fusion scheme (Waxman et al., 1995, 1996a,b,c), are applied to visual and thermal images of military relevant scenarios. The scenes represent 3 different scenarios that simulate military surveillance tasks. The images are registered around sunrise. At this time, the contrast in both image modalities is low. However, the visual images still provide a sufficient amount of detail to perceive the spatial structure of the scene. The thermal images clearly depict objects with large temperature contrast like persons, but they do not correctly represent the spatial context. The composite images produced by both fusion schemes clearly represent all details in their correct spatial context.
An observer experiment is performed to test if the increased amount of detail in the fused images can yield an improved observer performance in a task that requires a certain amount of situational awareness. The task that is devised involves the localization of a person in the displayed scene relative to some characteristic details that provide the spatial context. The results show that observers can indeed determine the relative location of a person in a scene with a significantly higher accuracy when they perform with fused images, compared to the individual image modalities. The MIT colour fusion scheme yields the best overall performance (i.e. an accuracy that is significantly higher than that obtained with images fused according to the TNO scheme and with the original images). Even the most simple (TNO) fusion scheme yields an observer performance that is better than that obtained for the individual (thermal and visual) images.

16. DESCRIPTORS	IDENTIFIERS
Target Acquisition Visual Displays	Colour Coding False Colour Image Fusion Multisensor Fusion Situational Awareness Spatial Localization

17a. SECURITY CLASSIFICATION (OF REPORT)	17b. SECURITY CLASSIFICATION (OF PAGE)	17c. SECURITY CLASSIFICATION (OF ABSTRACT)
18. DISTRIBUTION/AVAILABILITY STATEMENT Unlimited availability		17d. SECURITY CLASSIFICATION (OF TITLES)

Managementuittreksel

TNO Technische Menskunde, Soesterberg

titel	:	Beeldfusie verbetert situatie inschatting
auteurs	:	Dr. A. Toet, dr. J.K. IJspeert en M.J. van Dorresteijn
datum	:	31 oktober 1996
opdrachtnummer	:	A96/KL/347
IWP-nr.	:	786.1
rapportnr.	:	TM-96-A51

Deze studie werd uitgevoerd (*a*) om te onderzoeken onder welke omstandighden er winst te verwachten valt van de fusie van visuele beelden (CCD) en warmtebeelden (IR), en (*b*) of waarnemers onder dergelijke omstandigheden een situatie beter kunnen inschatten waaneer ze gefuseerde beelden gebruiken dan wanneer ze alleen over de oorspronkelijke beelden beschikken.

Beeldfusie is het combineren van beelden die afkomstig zijn van verschillende typen elektro-optische sensoren tot een enkel samengesteld beeld met een toegenomen informatieinhoud. Er wordt doorgaans aangenomen dat gefuseerde beelden de opname van informatie kunnen vereenvoudigen, mits gebruik wordt gemaakt van geschikte principes. Dit zou kunnen leiden tot een verbeterde taakprestatie wanneer de gebruiker in het gecombineerde beeld objecten beter kan detecteren en herkennen of nauwkeuriger en sneller kan lokaliseren. Tot nu toe is dit echter nooit onderzocht.

In dit onderzoek werden twee recent ontwikkelde false colour beeldfusietechnieken, de TNO methode (Toet & Walraven, 1996) en de MIT methode (Waxman *et al.*, 1995, 1996a,b,c), toegepast op visuele opnamen en warmtebeelden van militair relevante scenario's. De scenes representeren drie verschillende scenario's die militaire bewakingstaken simuleren. De beelden zijn rond zonsopgang opgenomen. Er is slechts weinig contrast in beide beeldmodaliteiten om die tijd. In de visuele beelden is echter voldoende detail te onderscheiden om de spatiële structuur van de scene te kunnen waarnemen. De thermische beelden representeren objecten met een hoog temperatuurcontrast, zoals personen, maar ze geven de spatiële context niet correct weer. De samengestelde beelden geproduceerd met beide fusie schema's geven alle details duidelijk weer in hun correcte spatiële context.

Er werd een waarnemingsexperiment uitgevoerd om vast te stellen of de toegenomen informatieinhoud van de gefuseerde beelden leidt tot een verbetering in de taakprestatie van waarnemers die een taak verrichten die een correcte inschatting van de situatie vereist. De taak komt neer op het localiseren van een persoon t.o.v. enkele karakteristieke details in de weergegeven scene.

Uit de resultaten blijkt dat waarnemers de relatieve positie van een persoon in een scene met een significant hogere nauwkeurigheid kunnen bepalen wanneer ze gebruik maken van gefuseerde beelden dan wanneer ze de originele beelden gebruiken. De MIT fusie methode levert de beste resultaten (d.w.z. een nauwkeurigheid die hoger is dan die wordt behaald met beelden die met de TNO methode zijn gefuseerd en dan die wordt behaald met de originele beelden). Zelfs de meest eenvoudige (TNO) fusie methode resulteert in een waarnemersprestatie die beter is dan die wordt behaald met de oorspronkelijke beelden.

CONTENTS

Report No.:	TM-96-A51
Title:	Image fusion improves situational awareness
Authors:	Dr. A. Toet, dr. J.K. IJspeert and M.J. van Dorresteijn
Institute:	TNO Human Factors Research Institute Group: Perception
Date:	December 1996
DO Assignment No.:	A96/KL/347
No. in Program of Work:	786.1

SUMMARY

Two recently developed false colour image fusion techniques, the TNO fusion scheme (Toet & Walraven, 1996) and the MIT fusion scheme (Waxman *et al.*, 1995, 1996a,b,c), are applied to visual and thermal images of military relevant scenario's. The scenes represent 3 different scenario's that simulate military surveillance tasks. The images are registered around sunrise. At this time, the contrast in both image modalities is low. However, the visual images still provide a sufficient amount of detail to perceive the spatial structure of the scene. The thermal images clearly depict objects with large temperature contrast like persons, but they do not correctly represent the spatial context. The composite images produced by both fusion schemes clearly represent all details in their correct spatial context.

An observer experiment is performed to test if the increased amount of detail in the fused images can yield an improved observer performance in a task that requires a certain amount of situational awareness. The task that is devised involves the localization of a person in the displayed scene relative to some characteristic details that provide the spatial context. The results show that observers can indeed determine the relative location of a person in a scene with a significantly higher accuracy when they perform with fused images, compared to the individual image modalities. The MIT colour fusion scheme yields the best overall performance (i.e. an accuracy that is significantly higher than that obtained with images fused according to the TNO scheme and with the original images). Even the most simple (TNO) fusion scheme yields an observer performance that is better than that obtained for the individual (thermal and visual) images.

Rap. nr. TM-96-A-51

TNO Technische Menskunde
Soesterberg

Beeldfusie verbetert situatie inschatting

A. Toet, J.K. IJspeert en M.J. van Dorresteijn

SAMENVATTING

Twee recent ontwikkelde false colour beeldfusietechnieken, de TNO methode (Toet & Walraven, 1996) en de MIT methode (Waxman *et al.*, 1995, 1996a,b,c), werden toegepast op visuele opnamen en warmtebeelden van militair relevante scenario's. De scenes representeren drie verschillende scenario's die militaire bewakingstaken simuleren. De beelden zijn rond zonsopgang opgenomen. Er is slechts weinig contrast in beide beeldmodaliteiten om die tijd. In de visuele beelden is echter voldoende detail te onderscheiden om de spatiele structuur van de scene te kunnen waarnemen. De thermische beelden representeren objecten met een hoog temperatuurcontrast, zoals personen, maar ze geven de spatiele context niet correct weer. De samengestelde beelden geproduceerd met beide fusie schema's geven alle details duidelijk weer in hun correcte spatiele context.

Er werd een waarnemingsexperiment uitgevoerd om vast te stellen of de toegenomen informatieinhoud van de gefuseerde beelden leidt tot een verbetering in de taakprestatie van waarnemers die een taak verrichten die een correcte inschatting van de situatie vereist. De taak komt neer op het localiseren van een persoon t.o.v. enkele karakteristieke details in de weergegeven scene.

Uit de resultaten blijkt dat waarnemers de relatieve positie van een persoon in een scene met een significant hogere nauwkeurigheid kunnen bepalen wanneer ze gebruik maken van gefuseerde beelden dan wanneer ze de originele beelden gebruiken. De MIT fusie methode levert de beste resultaten (d.w.z. een nauwkeurigheid die hoger is dan die wordt behaald met beelden die met de TNO methode zijn gefuseerd en dan die wordt behaald met de originele beelden). Zelfs de meest eenvoudige (TNO) fusie methode resulteert in een waarnemersprestatie die beter is dan die wordt behaald met de oorspronkelijke beelden.

1 INTRODUCTION

Scene analysis by a human operator may benefit from a combined or fused representation of images of the same scene taken in different spectral bands. For instance, after a period of extensive cooling (e.g. after a long period of rain or early in the morning) the visual bands may represent the background in great detail (vegetation or soil areas, texture), while the infrared bands are less detailed due to low thermal contrast in the scene. In this situation a target that is camouflaged for visual detection cannot be detected in the visual bands, but may be clearly represented in the infrared bands when it is warmer or cooler than its environment. The fusion of visible and thermal imagery on a single display may then allow both the detection and the unambiguous localization of the target (provided by the thermal image) with respect to the context (provided by the visual image).

The abovementioned line of reasoning is frequently adopted to promote image fusion, and has resulted in an increased interest in image fusion methods, as is reflected in a steadily growing number of publications on this topic (Fechner & Godlewski, 1995; Gove *et al.*, 1996; Li *et al.*, 1995; Waxman *et al.*, 1995, 1996a,b,c; Wilson *et al.*, 1995), and in the efforts of NATO AC/243, Panel 3, RSG.9 (e.g. NATO, 1993, 1994; Schwering, 1995; Sévigny, 1996).

A large effort has been spent on the development of new image fusion methods. However, until now there are no validation studies that investigate the applicability domain and the practical use of these techniques. Ultimately the performance of a fusion process must be measured as the degree to which it enhances a viewers ability to perform certain practical tasks. This study is performed (a) to investigate the conditions for which the fusion of visual and thermal images may result in a single composite image with extended information content, and (b) to test the capability of two recently developed colour image fusion schemes (Toet & Walraven, 1996; Waxman *et al.*, 1995, 1996a,b,c) to enhance the situational awareness of observers operating under these specific conditions and using visual and thermal images.

2 METHODS

2.1 Image registration

Apparatus

The *visible-light* camera was a Siemens K235 Charge Coupled Device (CCD) videocamera, equipped with a remotely controlled COSMICAR C10ZAME-2 (Asahi Precision Co. Ltd., Japan) zoom lens (f = 10.5–105 mm; 1:1,4). The *infrared* (IR) camera was an Amber Radiance 1 (Goleta, CA, U.S.) Focal Plane Array (FPA)-camera, operating in the 3–5 μm

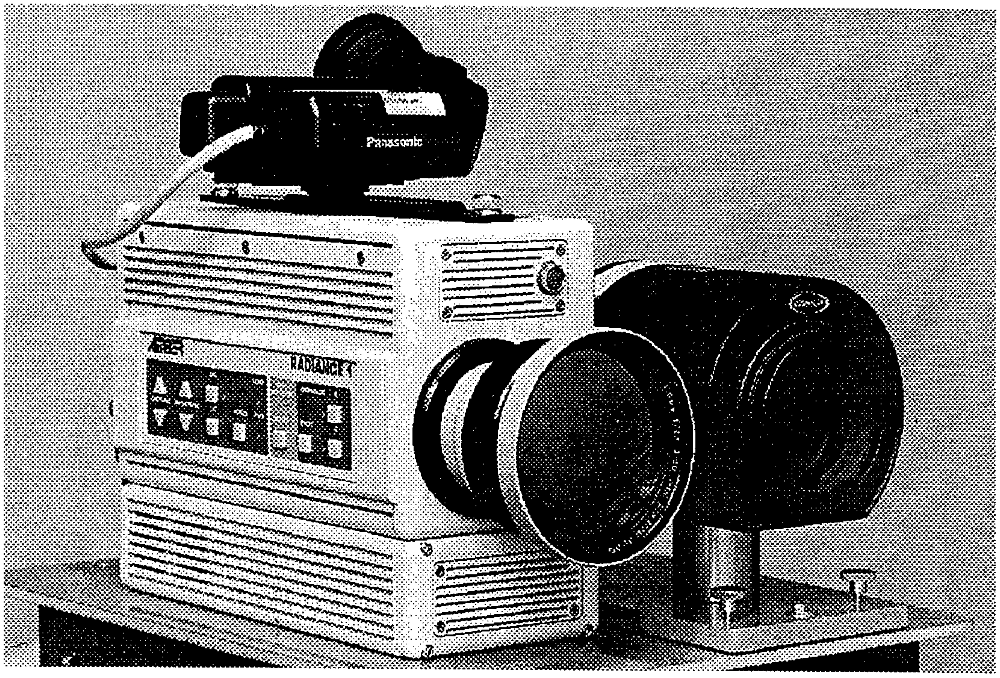


Fig. 1 The multispectral imaging sensor suite, consisting of an Amber Radiance 1 IR camera (left) and a Siemens K235 CCD videocamera (right).

(mid-range) band, and equipped with a 100 mm f/2.3 Si-Ge-lens. Each pixel corresponds to a square 1.3 min of arc wide instantaneous field of view. Since the entire array consists of 256 × 256 pixels, the total field of view is about 5.6 degrees wide.

The CCD camera and the IR camera are mounted together on a single, tripod supported platform.

The CCD and IR images must be aligned before they can be fused. The signals of the thermal and visual cameras are therefore spatially registered as closely as possible. This is effected by (a) chosing a set of fiducial points that are visible both in the thermal and in the visual images and (b) ensuring that their positions coincide as much as possible at each location over the common field of view of both cameras. Well defined fiducial points are created by placing 3 large plastic jerrycans in the scene. The cans are bright (white) and filled with hot water so that they are clearly visible in both image modalities. They are evenly distributed over a horizontal row such that the outer two are just inside the field of view. The difference signal of the CCD and IR cameras is displayed on a CRT. Using both this difference image and the fiducial points both the zoom ratio (display size of the common field of view) and the optical axis of the CCD camera are adjusted to those of the IR camera. This tuning is iteratively performed with the fiducial points at different positions in the common field of view (for different orientations of the optical axis of the camera suite). When the optimal horizontal image registration has been achieved there is still a size difference of about 5% in the vertical direction. A final correction for this misalignment is performed by geometrically transforming and resampling the digitized video frames. By registering the scene for 3 different vertical positions (top, middle, bottom) of the virtual line joining the 3 jerrycans (situated in the common field of view of both cameras) 9 matching points are obtained. For the scenes used in this experiment an additional number of matching points is opportunistically available, because a row of street lamps is clearly visible from the viewing location. The field of view of the cameras covers a highway section containing 6 lamps. By registering this highway scene for 3 different vertical positions of the row of lamps an additional number of 18 matching points is obtained. The scenes containing the matching sets of fiducial points are registered at the beginning of each recording session. Digitized frames from this recording are used to compute the coefficients of the affine warping transformation that maps corresponding points in the scene to corresponding pixel locations in the image plane. The digitized CCD images are then warped and resampled so that their pixels are in registration with the corresponding pixels of the IR image. The alignment of both cameras is not changed during a session. Therefore a single set of match points can be used to register all scenes that are recorded during a session.

The camera signals are stored on video tape using two high fidelity Panasonic AG-6730E SVHS tape recorders. Individual frames are time stamped using a bar code generator.

2.2 Conditions

Visual and thermal images are registered at different times of the day for a period of 3 weeks. The weather conditions range from partially clouded to rain and fog. Most of the time all details that can be seen in the visible image are also represented in the thermal image.

The CCD and IR images that are selected for the observer validation study are registered at The Hague, The Netherlands, from 24–25 September 1996, and between 07:30 and 08:30 hours MET in the morning. During the recording period the contrast in both image modalities is low. The visual contrast is low because the recording period is just before sunrise. The thermal contrast is low because most of the objects in the scene have about the same temperature after having lost their excess heat by radiation during the night.

The set-up was located in a room at the 10th floor of the TNO-FEL building. This location was chosen to simulate (i) surveillance cameras mounted on high poles, and (ii) the view from a helicopter during a low nap-of-the-earth flight. The room was not heated. The sensor suite was placed near the windows, which were open when the system was recording (glass is opaque for the IR camera). The signals of both cameras are degraded because water vapour condensed inside the sensor systems. As a result a large number of pixels near the corners of the IR image drop out and a horseshoe shaped blemish appears in the mid-lower part of the CCD image. This is not problematic for the present study since the region of interest is not close to the image imperfections anyway.

2.3 Image fusion

The images were fused using two recently developed color image fusion schemes.

The MIT fusion scheme

The computational image fusion scheme developed at MIT Lincoln Labs (Waxman *et al.*, 1995, 1996a,b,c) derives from biological models of color vision and fusion of visible light and infrared (IR) radiation.

In the case of color vision in monkeys and man, retinal cone sensitivities are broad and overlapping, but the images are quickly contrast enhanced *within bands* by spatial opponent processing (via cone-horizontal-bipolar cell interactions) creating both ON and OFF center-surround responsee channels (Schiller, 1992). These signals are then contrast enhanced *between* bands via interactions among bipolar, sustained amacrine, and single-opponent color ganglion cells (Schiller & Logothetis, 1990; Gouras, 1991), all within the retina.

Fusion of visible and thermal IR imagery has been observed in the optic tectum of rattlesnakes and pythons (Newman & Hartline, 1981, 1982). These neurons display interactions in which one modality (e.g. IR) can enhance or depress the response to the other sensing modality (e.g. visible) in a strongly nonlinear fashion.

Spectral reflectivity ρ and emissivity ϵ are linearly related at each wavelength λ: $\rho(\lambda) = 1 - \epsilon(\lambda)$. This provides a further argument for choosing an opponent scheme to fuse visible and thermal imagery.

In the color image fusion scheme the individual input images are first enhanced by filtering them with a feedforward center-surround shunting neural network (Grossberg, 1988). This operation serves to (i) enhance spatial contrast in the individual visible and IR bands, (ii) to create both positive and negative polarity IR contrast images, and (iii) to create two types of single-opponent color contrast images. The resulting single-opponent color contrast images represent grayscale fused images that are analogous to the IR-depressed visual and IR-enhanced visual cells of the rattlesnake (Newman & Hartline, 1981, 1982). To obtain a natural color representation of these single-opponent images (each being an 8-bit grayscale image) the enhanced visual image is assigned to the green channel, the difference signal of respectively the enhanced visual and infrared images is assigned to the blue channel, and the sum of the enhanced visual and infrared images is assigned to the red channel of an RGB display:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} \mathrm{CCD}^+ & + & \mathrm{IR}^+ \\ \mathrm{CCD}^+ & & \\ \mathrm{CCD}^+ & - & \mathrm{IR}^+ \end{pmatrix} , \qquad (1)$$

where the $()^+$ indicates the enhanced version. These channels correspond with our natural associations of warm (red) and cool (blue). The result is a fused color representation of visible and infrared imagery.

The TNO fusion scheme

TNO (Toet and Walraven, 1996) recently presented a color image fusion scheme that can be regarded as a simple approximation of the more refined MIT scheme.

First, the common component of the two original input images is determined. This is simply implemented as a local minimum operator. The common component of two images $I_1(i,j)$ and $I_2(i,j)$ is therefore given by

$$I_1 \cap I_2\ (i,j)\ =\ \text{Min}\ \{I_1(i,j), I_2(i,j)\}\ , \tag{2}$$

where I_1 and I_2 represent sampled 2D luminance functions or digital images, and i,j represent the indices of an element of the sampling grid (the pixel coordinates; $\{i,j\} \in [1,256]$, $\{I_1, I_2\} \in [0,255]$).

Second, the common component is subtracted from the original images to obtain the unique or characteristic components I_1^* and I_2^* of the two registered images $I_1(i,j)$ and $I_2(i,j)$:

$$I_1^*\ =\ I_1\ -\ I_1 \cap I_2\ ,\ \text{and} \tag{3}$$

$$I_2^*\ =\ I_2\ -\ I_1 \cap I_2\ . \tag{4}$$

These images represent the details that are unique to the corresponding image modalities.

Third, the unique component of each image modality is subtracted from the image of the other modality:

$$I_3\ =\ (I_1 - I_2^*)\ \bowtie\ (I_2 - I_1^*)\ \ , \tag{5}$$

where $\bowtie$ represents the fusion operator, which represents any operation that effectively combines information from two input images into a single output image. The subtraction operation has been arbitrarily chosen. Any operation that reduces the dynamic range of one image at locations where the characteristic component of the other image has an appreciable value can in principle be used. This step is similar to the aforementioned color-opponency found in biological color vision. It serves to enhance the representation of sensor specific details in the final fused result.

Finally, a fused color image I_3 is produced by displaying the images resulting from step 3 through respectively the red and green channels of a color display. In this study, I_1 corresponds to the processed thermal image and I_2 to the processed video image, so that I_3 is given by:

$$I_3 = \begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} \text{IR} & - & \text{CCD}^* \\ \text{CCD} & - & \text{IR}^* \\ & 0 & \end{pmatrix}\ . \tag{6}$$

The resulting color rendering enhances the visibility of certain details and preserves the specificity of the sensor information. The fused images also have a fairly natural appearance.

The fusion scheme involves only simple operations. The method can therefore probably be applied in real time and with relatively simple hardware.

The algorithm operates only on corresponding pixel values. As a result, it does not degrade the resolution of the final result. Standard image processing techniques can be used to enhace image contrast or extract image features prior to the fusion process. However, in the present study the images were not processed before they were fused.

2.4 Stimuli

The stimuli used in this experiment are 6 different types of images:

- graylevel images representing the signal of the video (CCD) camera,
- graylevel images representing the signal of the infrared (IR) camera,
- color images representing the result of the fusion of corresponding CCD and IR image pairs (i.e. the combination of CCD and IR images of the same scene and registered at the same instant) using the MIT scheme,
- graylevel images representing luminance component of the abovementioned color fused images,
- color images representing the result of the fusion of corresponding CCD and IR image pairs using the TNO scheme, and
- graylevel images representing luminance component of the abovementioned color fused images.

The graylevel images are quantized to 8 bits. The color images are quantized to 24 bits (8 bits for each of the RGB channels).

The individual images correspond to successive frames of a time sequence. The time sequences represent 3 different scenarios. These scenarios were developed by the Royal Dutch Army (Buimer, 1993) They simulate surveillance tasks and were chosen because of their military relevance.

Scenario I: guarding a UN camp

This scenario corresponds to a typical and statical peace keeping operation, for instance guarding of a UN camp (Buimer, 1993: Scenario 1). The sector that is monitored corresponds to the corner of a fence that encloses a military asset. In the original scenario, the assignment of the observer (guard) is to detect infiltration attempts and hostile actions of subversive elements in a very early stage. To distinguish innocent bypassers from individuals planning to perform subversive actions the observer needs to know at each moment the position of the individual relative to the fence. Persons showing an unusual interest or loitering in certain critical locations are a priori suspect. The operator must therefore be able to determine the exact position of a person in the scene at any time.

During the registration period the fence is clearly visible in the CCD image. In the IR image however, the fence is merely represented by a vague haze. A person (walking along the fence) is clearly visible in the IR image but can hardly be distinguished in the CCD image. In the fused images both the fence and the person are clearly visible. An observer's situational awareness can therefore be tested by asking the subject to report the position of the person relative to the fence.

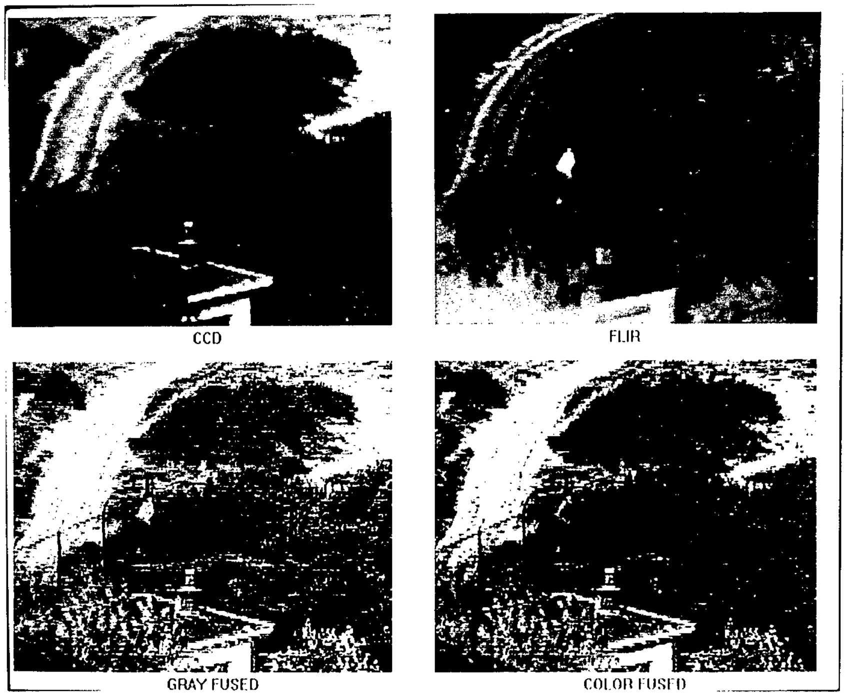


Fig. 2 Original CCD (upper left), IR (upper right), MIT graylevel fused (lower left), and MIT color fused (lower right) images of Scenario I.

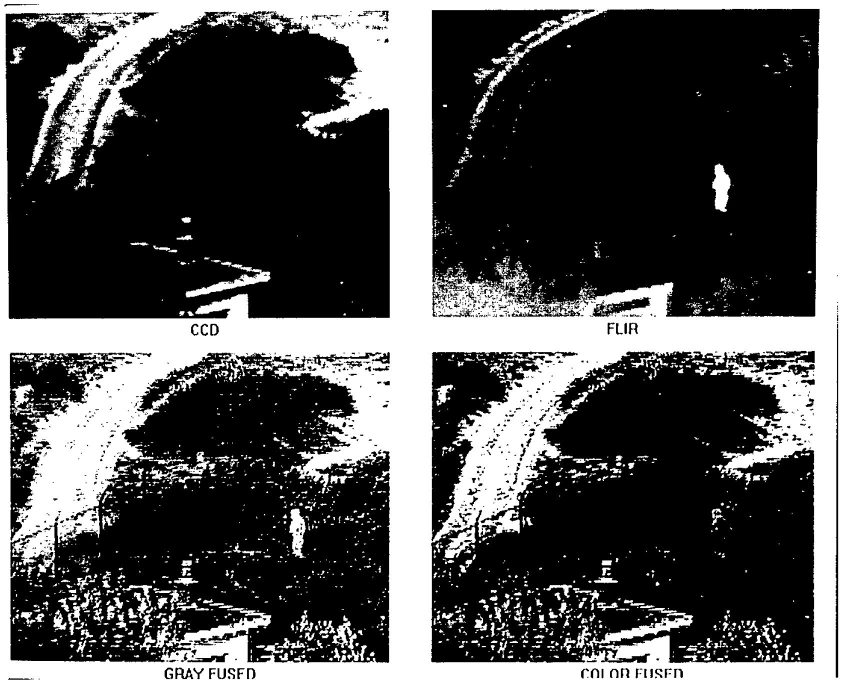


Fig. 3 As Fig. 2 for a different position of the person in the scene.

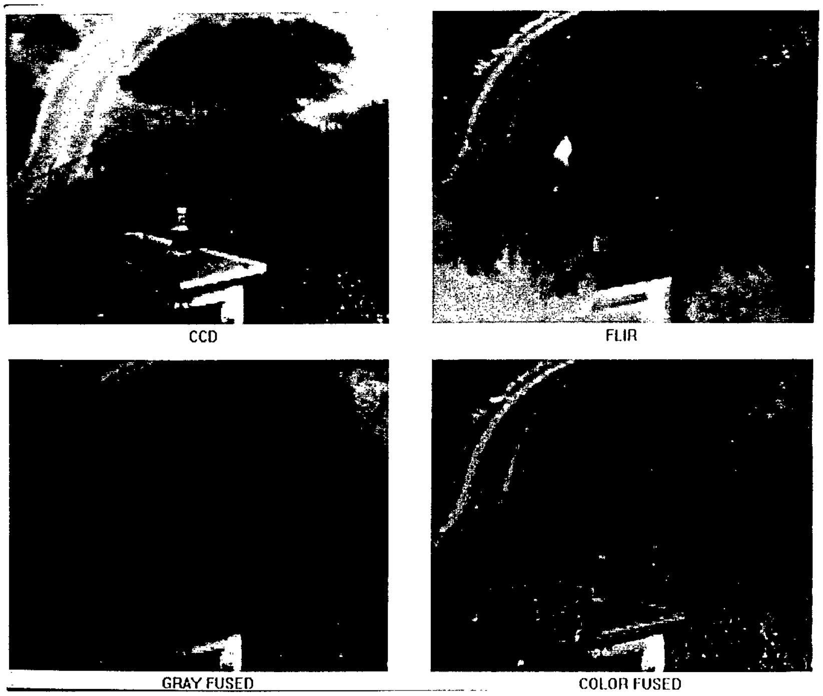


Fig. 4 Original CCD (upper left), IR (upper right), TNO graylevel fused (lower left), and TNO color fused (lower right) images of Scenario I.

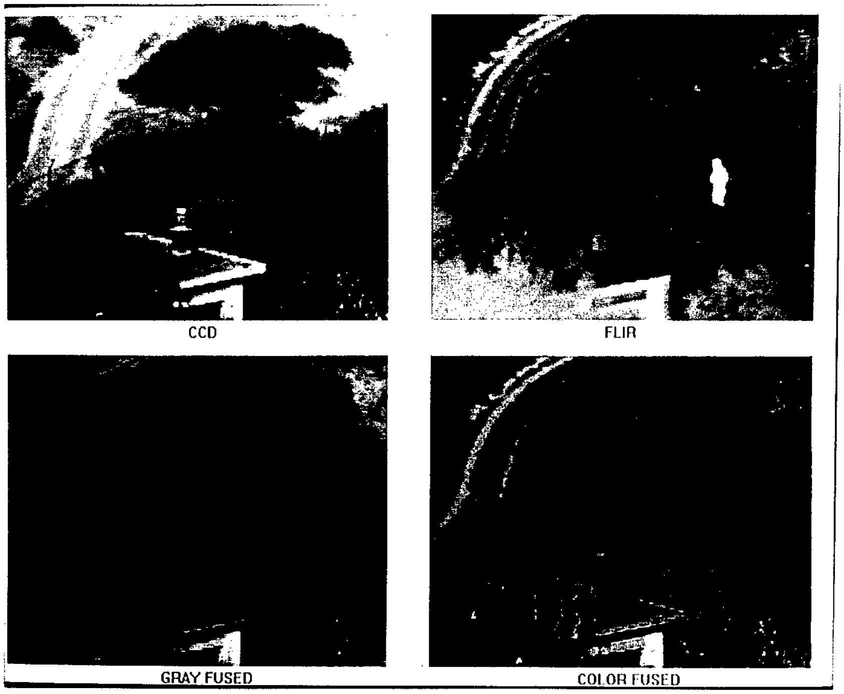


Fig. 5 As Fig. 4 for a different position of the person in the scene.

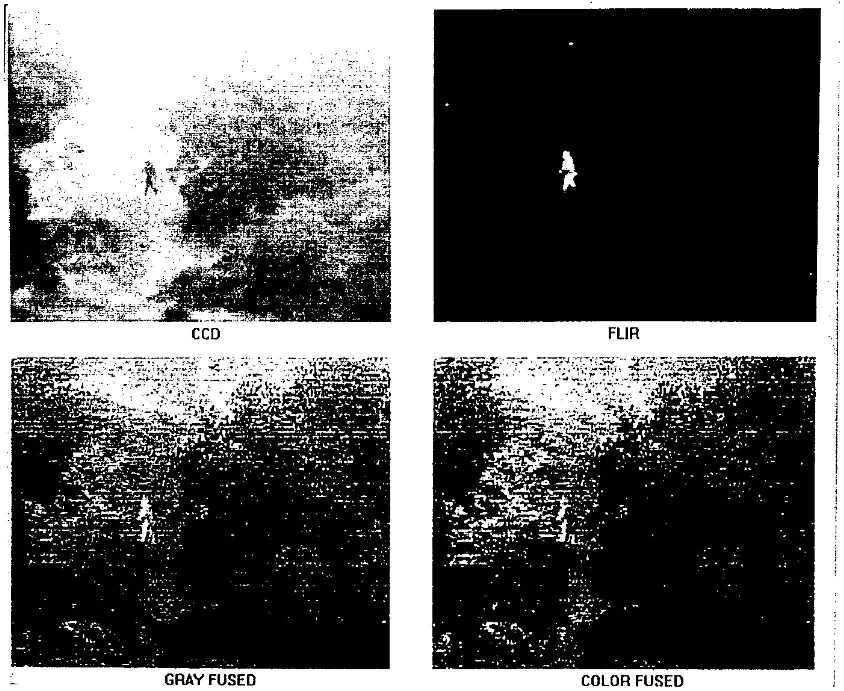


Fig. 6 Original CCD (upper left), IR (upper right), MIT graylevel fused (lower left), and MIT color fused (lower right) images of Scenario II.

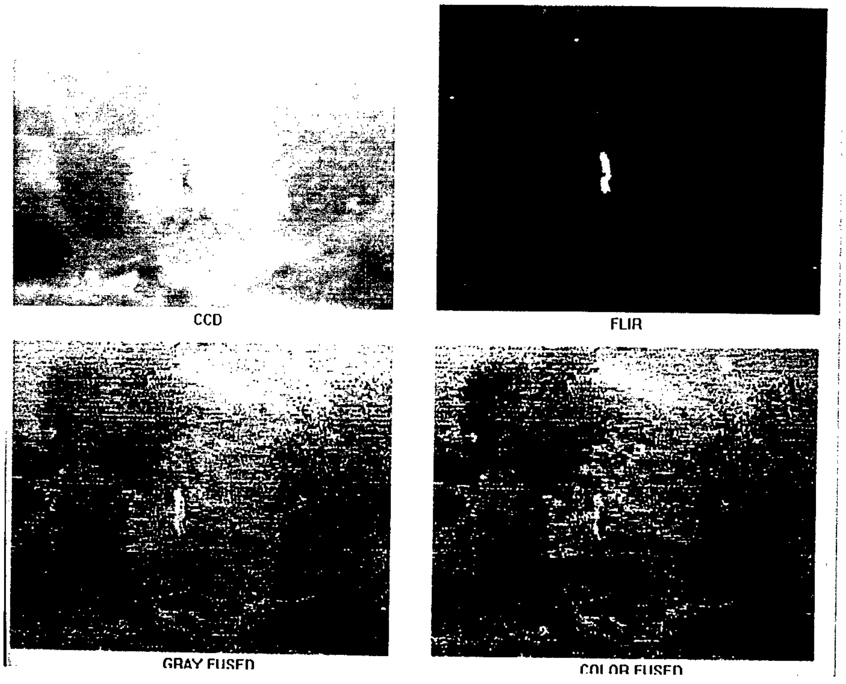


Fig. 7 As Fig. 6 for a different position of the person in the scene.

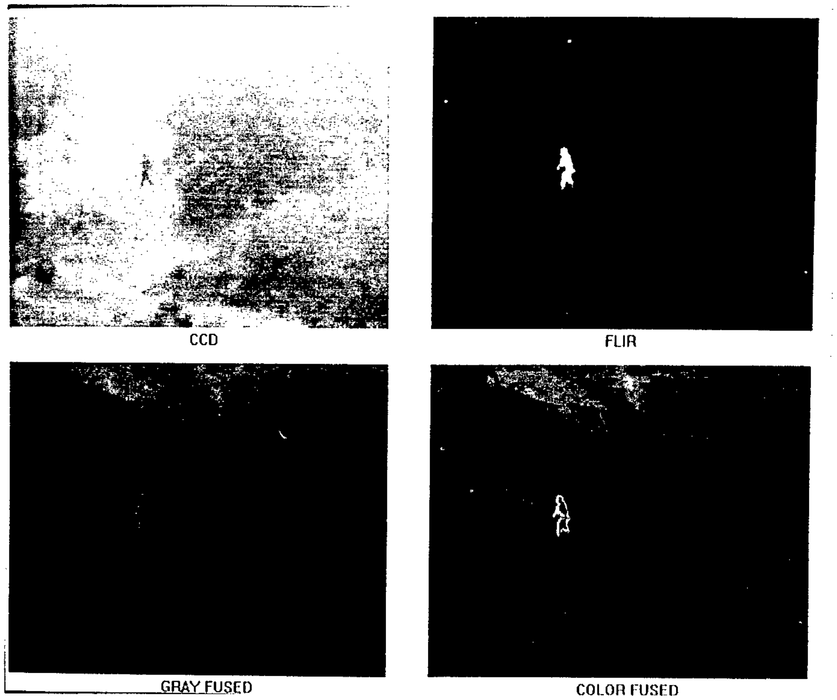


Fig. 8 Original CCD (upper left), IR (upper right), TNO graylevel fused (lower left), and TNO color fused (lower right) images of Scenario II.

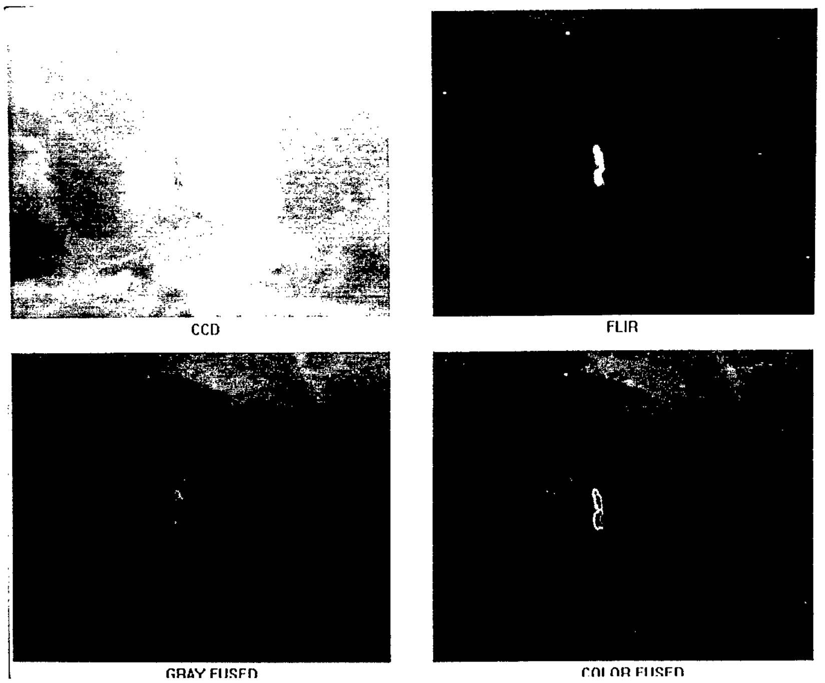


Fig. 9 As Fig. 8 for a different position of the person in the scene.

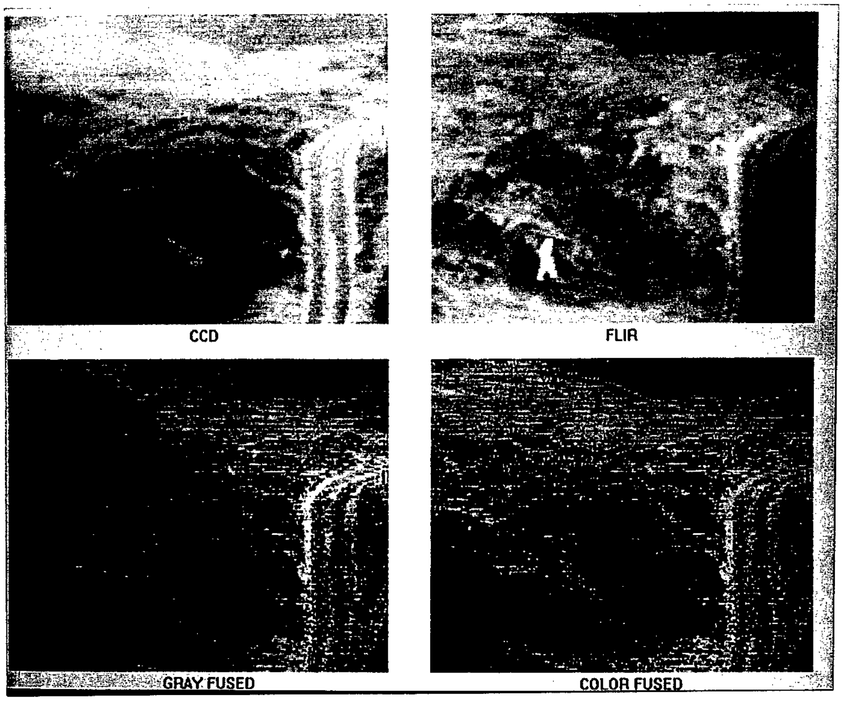


Fig. 10 Original CCD (upper left), IR (upper right), MIT graylevel fused (lower left), and MIT color fused (lower right) images of Scenario III.

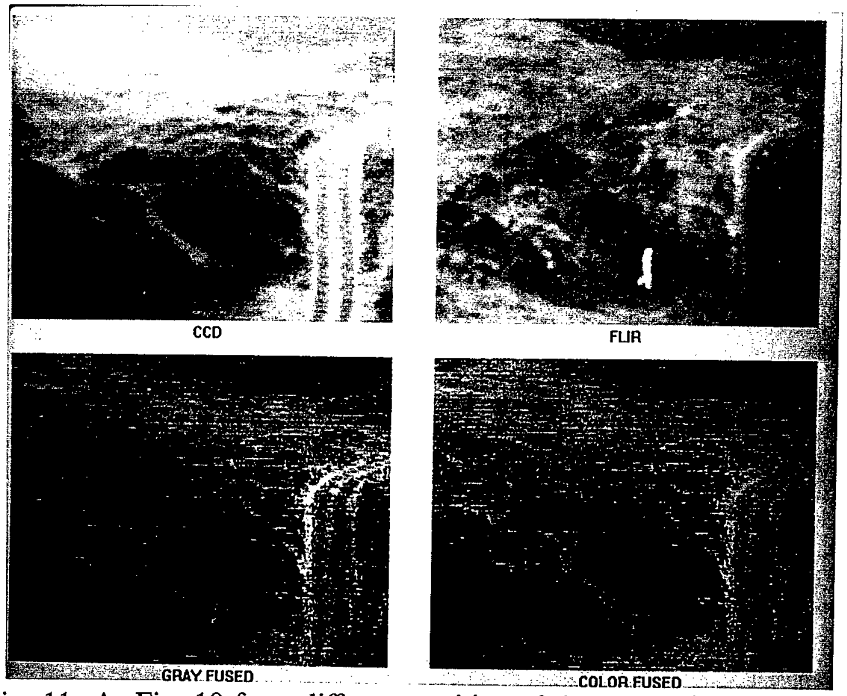


Fig. 11 As Fig. 10 for a different position of the person in the scene.

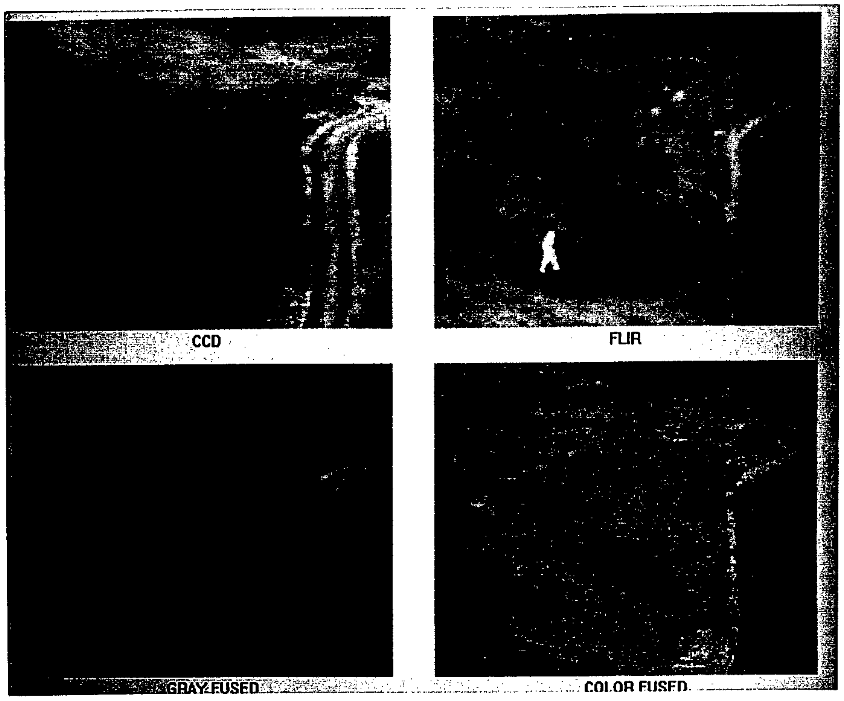


Fig. 12 Original CCD (upper left), IR (upper right), TNO graylevel fused (lower left), and TNO color fused (lower right) images of Scenario III.

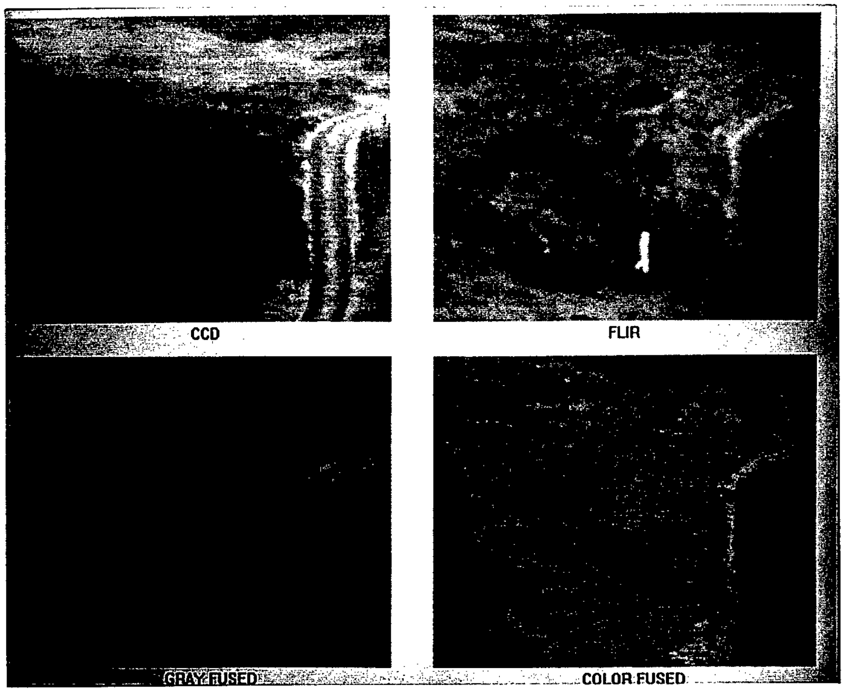


Fig. 13 As Fig. 12 for a different position of the person in the scene.

Scenario II: guarding a temporary base

This scenario corresponds to a typical and statical peace enforcing operation, for instance guarding a temporary base that can be used for offensive actions (Buimer, 1993: Scenario 4). The narrow sector that is monitored corresponds to a likely infiltration route through the terrain surrounding the base. The terrain is dune like. Only a small section of the terrain is visible, the rest is occluded by trees. In the original scenario the assignment of the observer (guard) is similar to that of Scenario I: infiltration attempts and hostile actions of subversive elements that will try to destroy military goods must be detected and countered in a very early stage.

During the registration period the trees appear larger in the IR image than they really are because they have nearly the same temperature as their local backgound. In the CCD image however, the contours of the trees are correctly represented. A person (crossing the interval between the trees) is clearly visible in the IR image but is represented with low contrast in the CCD image. In the fused images both the outlines of the trees and the person are clearly visible. As a result it is difficult to determine the position of the person relative to the trees using either the CCD or the IR images. The fused images correctly represent both the contours of the trees and the person. An observer's situational awareness can therefore be tested by asking the subject to report the position of the person relative to the midpoint of the interval delineated by the contours of the trees that are positioned on both sides of the person.

Scenario III: guarding a large area

This scenario also corresponds to a statical peace enforcing operation, for instance the surveillance of a large area (Buimer, 1993: Scenario 5). The scene represents a dune landscape, covered with semi-shrubs and sandy paths. In the original scenario the assignment of the observer (guard) is to detect any attempt to infiltrate a certain area.

During the registration period the sandy paths in the dune area have nearly the same temperature as their local backgound, and are therefore represented with very low contrast in the IR image. In the CCD image however, the paths are depicted with high contrast. A person (walking along a trajectory that intersects the sandy path) is clearly visible in the IR image but is represented with less contrast in the CCD image. In the fused images both the outlines of the paths and the person are clearly visible. It is difficult (or even impossible) to determine the position of the person relative to the sandy path he is crossing from either the IR or the CCD images. An observer's situational awareness can therefore be tested by asking the subject to report the position of the person relative to the sandy path.

2.5 Apparatus

A Pentium 100 MHz computer, equipped with a Diamond SVGA board, is used to present the stimuli, measure the response times and collect the observer responses. The stimuli are presented on a 17 inch Vision Master (Iiyama Electric Co., Ltd) color monitor, using the 640 × 480 pixels mode and a 100 Hz refresh rate.

2.6 Procedure

The subject's task is to assess from each presented image the position of the person in the scene relative to some characteristic terrain features. The reference features are clearly represented in the visual images, but not shown in the thermal images. The person is most clearly represented in the thermal image, but can hardly be perceived in the visible images.

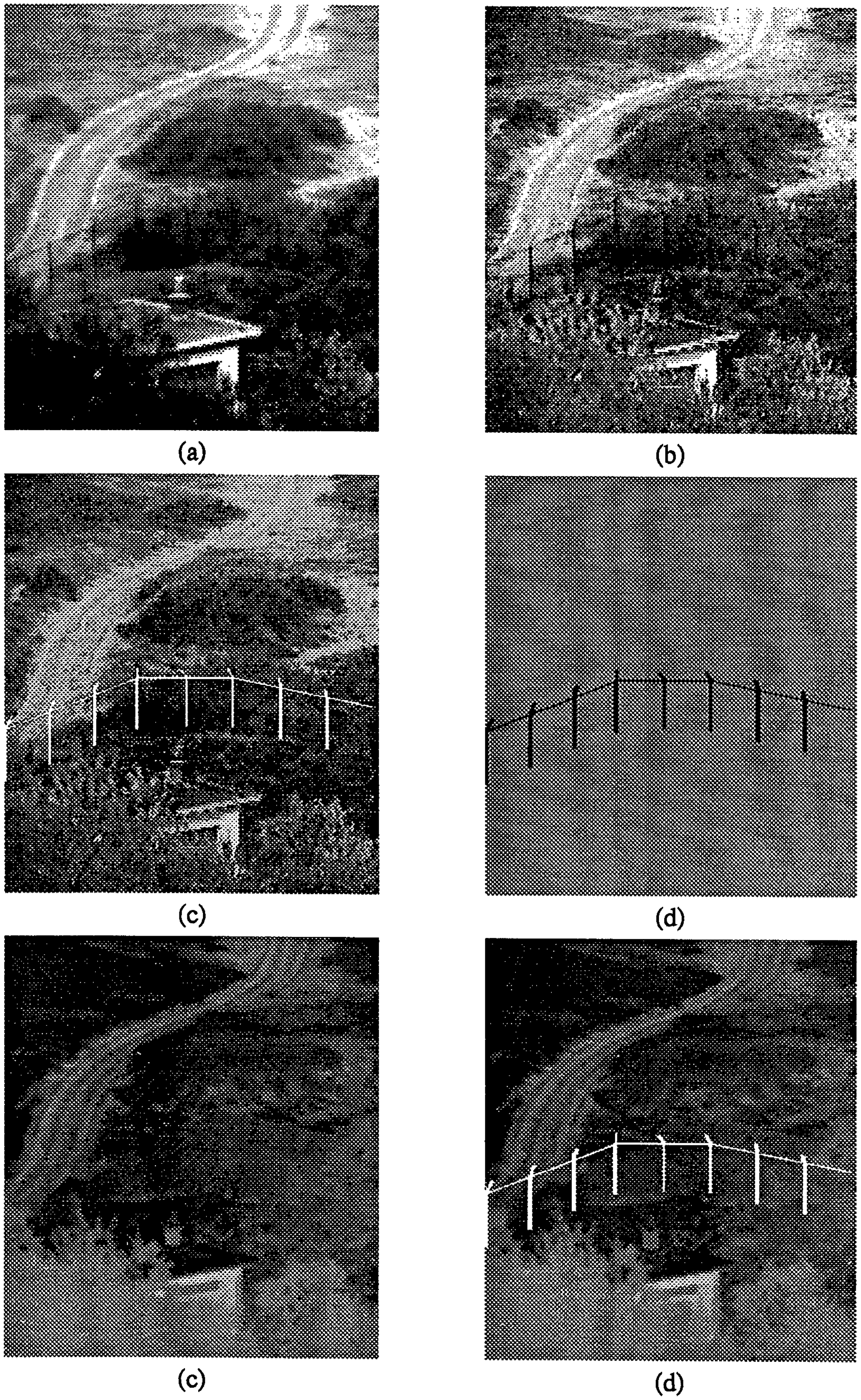
In Scenario I the reference features are the poles that support the fence. These poles are clearly visible in the CCD images, but not represented in the IR images because of they have almost the same temperature as the surrounding terrain. The person that walks along the fence is most of the time not represented in the CCD images, but is represented at high contrast in the thermal images. A total of 9 frames with the person at different locations is used in the experiment. The different locations of the person in these images are equally distributed along the entire length of the visible part of the fence (the reference interval).

In Scenario II the outlines of the trees serve to delineate the reference interval. The contours of the trees are correctly represented in the CCD images. However, in the IR images the trees appear larger than their physical size because they almost have the same temperature as the surrounding soil. As a result, the scene is incorrectly segmented after quantization and it is not possible to perceive the correct borders of the area between the trees. The person that walks between the trees is represented at low contrast in the CCD images, and at high contrast in the thermal images. A total of 9 frames with the person at different locations is used in the experiment. The different locations of the person in these images are equally distributed over the entire gap between the trees (the reference interval).

In Scenario III the area of the small and winding sandy path provides a reference contour for the task at hand. This path is represented at high contrast in the CCD images, but it is not represented in the IR images because it has the same temperature as the surrounding soil. A person walks along a trajectory that intersects this path. This person is represented at very low contrast in the CCD images, and at high contrast in the thermal images. A total of 9 frames with the person at different locations is used in the experiment. The different locations of the person in these images are equally distributed over both sides of the sandy path (the reference feature).

Each stimulus is presented for 1 s. The position of the center of the image is randomly jittered around the center of the screen between presentations. The horizontal and vertical integer pixel offset coordinates $\Delta x, \Delta y$ are randomly selected integers such that $\Delta x \in [-20, 20]$ and $\Delta y \in [-40, 40]$. This position jitter is introduced to ensure that subjects can not refer to previously perceived scenes to make their judgement.

A schematical representation of the reference features is shown immediately after each stimulus presentation. The subject's task is to indicate the perceived location of the person in the scene by placing a mouse controlled cursor at the corresponding location in



(a)

(b)

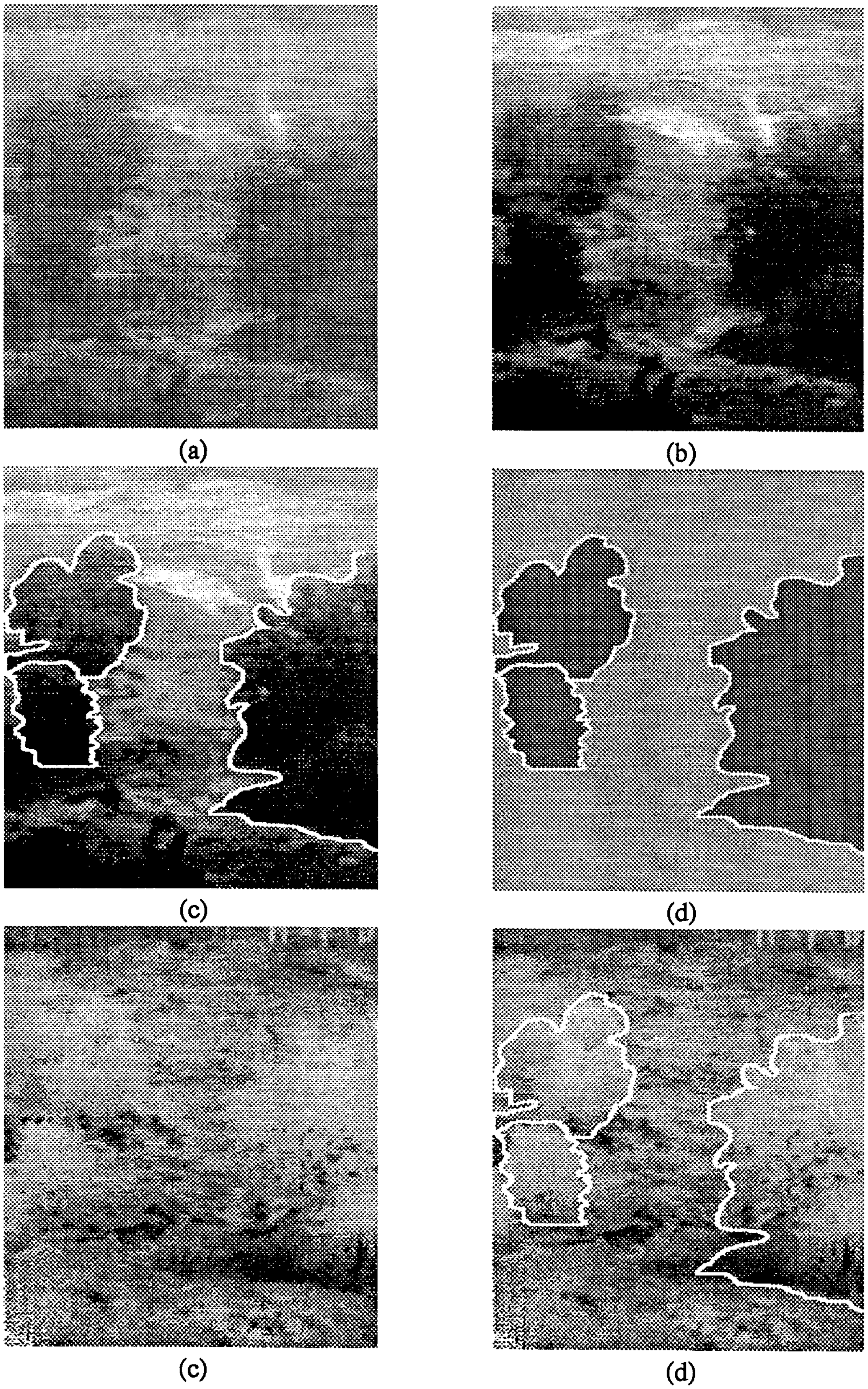
(c)

(d)

(c)

(d)

Fig. 14 The construction of the reference image for Scenario I. (a) The original CCD image, (b) its contrast enhanced version, (c) the outlines of the fence drawn over the enhanced CCD image, (d) the binary reference image that is used in the experiments, (e) the original corresponding IR image, and (f) the outlines of the fence drawn over the IR image.



(a)

(b)

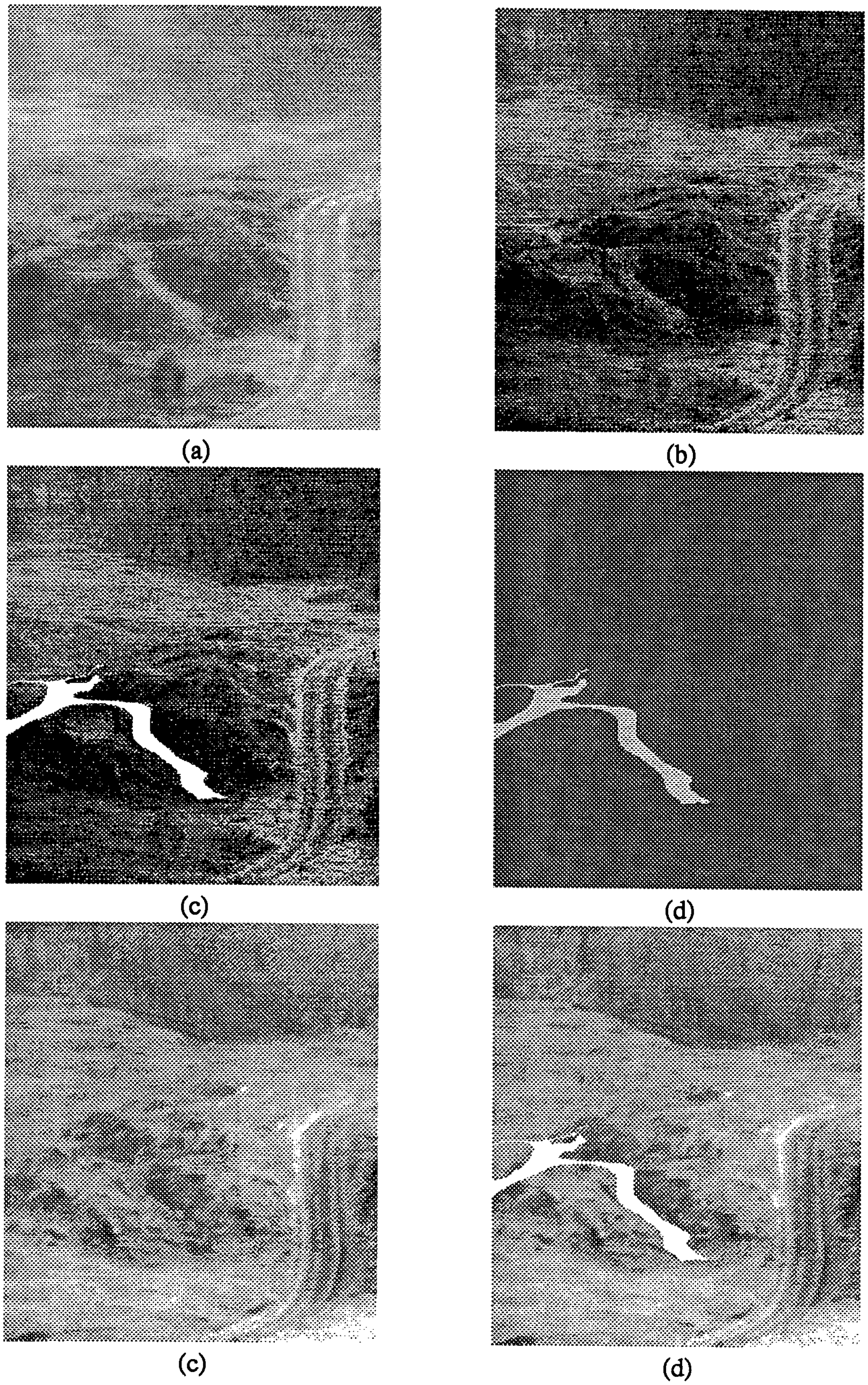
(c)

(d)

(c)

(d)

Fig. 15 As Figure 14 for Scenario II. In this case the reference contours are defined by the outlines of the trees.



(a)

(b)

(c)

(d)

(c)

(d)

Fig. 16 As Figure 14 for Scenario III. In this case the reference feature is the sandy path.

this schematical drawing. The subject has in principle unlimited time to reach a decision. When the left mouse button is pressed the computer registers the coordinates corresponding to the indicated image location (the mouse coordinates) and computes a weighted distance in the image plane between the actual position of the person and the indicated location. The actual position of the person on each image is stored on file. The subject's task mainly involves the localization of an upright person with respect to vertically elongated image contours. Therefore an elliptical distance function is adopted that attributes horizontal distances a weight which is 4 times larger than the weight factor corresponding to equal vertical distances. The subject presses the right mouse button if the person in the displayed scene has not been detected.

A complete run consists of 162 presentations (6 image modalities × 3 scenario's × 9 frames per scenario), and typically lasts about one hour.

The schematical representations are constructed from the original images by

- applying standard image processing techniques like histogram equalization and contrast stretching to enhance the representation of the reference contours in the visual images,
- drawing the contours of the reference features (judged by eye) on a graphical overlay on the contrast enhanced visual images, and
- filling the contours with a homogeneous graylevel value, using the Adobe Photoshop 3.1 image processing package.

The images thus created represent a segmented version of the reference features and map one to one to the original images. The subject can only perform the localization task by memorizing the perceived position of the person relative to the reference features.

2.7 Subjects

A total of 6 subjects, aged between 20 and 30 years, serve in the experiments reported below. All subjects have corrected to normal vision, and no known color deficiencies.

2.8 Viewing conditions

Viewing is binocular. The experiments are performed in a dimly lit room. The images are projected onto the screen of a CRT display. This screen subtends a viewing angle of 25.5 × 19.5 degrees at a viewing distance of 0.60 m.

3 RESULTS

Figure 17 shows the mean weighted distance between the actual position of the person in each scene and the position indicated by the subjects (the perceived position), for each of the 4 images modalities and for both the TNO and the MIT fusion schemes. A low value of this mean weighted distance measure corresponds to a high observer accuracy and a correctly perceived position of the person in the displayed scenes relative to the main reference features. High values correspond to a large discrepancy between the perceived position and the actual position of the person. In all scenarios the person was at approximately 300 m distance from the viewing location. At this distance one pixel corresponds to 11.4 cm in the field.

Figure 17 shows that the observers achieve the best overall performance in the relative spatial localization task with the images that are fused according to the MIT scheme. The images fused according to the TNO scheme yield an accuracy that is somewhat lower than the accuracy that is observed with the MIT scheme (p=0.017). Performance accuracy with the individual thermal and visual image modalities is significantly lower than the accuracy obtained with composite images produced by the two fusion schemes (p=0.0023 for the TNO scheme and p=0.0021 for the MIT scheme). The lowest accuracy is achieved for the thermal images. The visual images appear to yield a slightly higher accuracy. However, this accuracy is misleading since observers do not detect the person in a large percentage of the visual images, as shown by Figure 18. The difference between the results for the graylevel fused and the colour fused images is not significant for both fusion schemes (p=0.134 for the MIT scheme and p=0.398 for the TNO scheme).

Figure 18 shows that the observers do not know the location of the person in the scene for 20% of the visual images. For the thermal image this number is 8%. The graylevel fused images result in a smaller fraction of "don't know" replies (respectively 4.5% for the MIT grayfused images and 4.9% for the TNO results). The lowest number of misses is obtained for the colour fused images (respectively 1.5% for the MIT grayfused images and 1.9% for the TNO results).

4 DISCUSSION

This study investigates (*a*) for which conditions the fusion of visual and thermal images results in a single composite image with extended information content, and (*b*) whether two recently developed colour image fusion schemes (Toet & Walraven, 1996; Waxman *et al.*, 1995, 1996a,b,c) can enhance the situational awareness of observers operating under these specific conditions and using visual and thermal images.

Conditions in which fusion of visual and thermal imagery is most likely to result in images with increased information content occur around sunrise. At this time the contrast of both the visual and the thermal images is very low.

The visual contrast is low around sunrise because of the low luminance of the sky. However, contours of extended objects are still visible. After some image enhancement (like histogram equalization or contrast stretching) even an appreciable amount of detail can be perceived. Small objects with low reflectance, like a person wearing a dark suit or camouflage clothing, or objects that are partly obscured, are not represented in the visual image under these conditions, and can therefore not be detected.

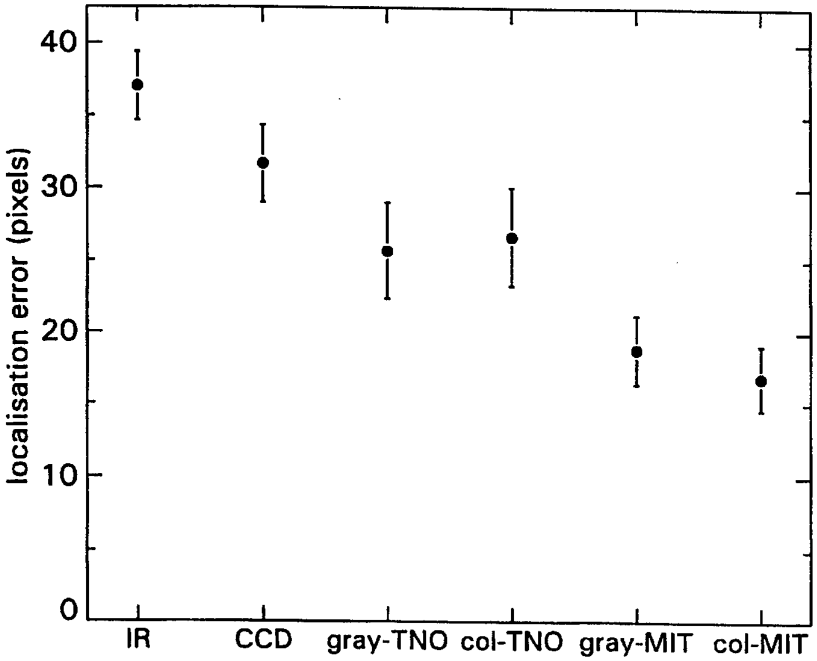


Fig. 17 The mean weighted distance between the actual position of the person in each scene and the perceived position for each of the 4 images modalities, for both the MIT and the TNO fusion scheme. The error bars indicate the size of the standard error in the perceived location.

The thermal contrast is low around sunrise because most of the objects in the scene have about the same temperature after losing their excess heat by radiation during the night. As a result the contours of extended objects are not at all or incorrectly represented in the thermal image.

The fusion of images registered around sunrise should therefore result in images that represent both the context (the outlines of extended objects) and the details with a large thermal contrast (like persons) in a single composite image.

To test this hypothesis a large set of image sequences is registered around sunrise on different days. The scenes used in this study represent 3 different scenarios that were developed by the Royal Dutch Army (Buimer, 1993) The images are fused using both the TNO and the MIT colour fusion schemes. Graylevel fused images are also produced by taking the luminance component of both types of colour fused images. The results show that the fusion of thermal and visual images indeed results in composite images with an increased amount of detail.

An observer experiment is performed to test if the increased amount of detail in the fused images can yield an improved observer performance in a task that requires a certain amount of situational awareness. The task that is devised involves the localization of a person in the displayed scene relative to some characteristic details that provide the spatial context.

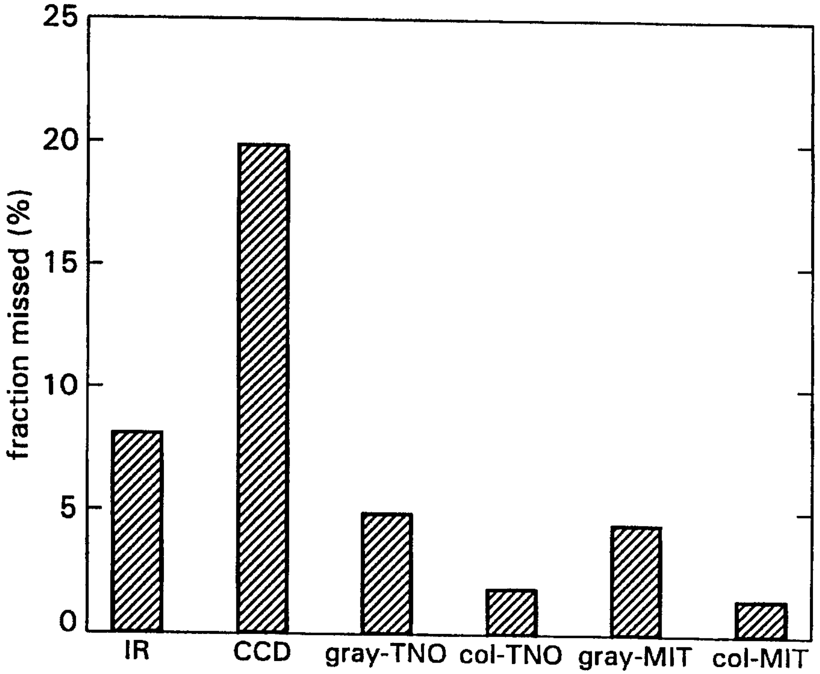


Fig. 18 The percentage of image presentations in which the person is not detected by the observer.

The person is optimally represented in the thermal imagery and the reference features are better represented in the visual imagery. The hypothesis is therefore that the fused images provide a better representation of the spatial structure of the depicted scene.

To test this hypothesis subjects perform a relative spatial localization task with a selection of thermal, visual, and (both graylevel- and colour-) fused images representing the abovementioned military scenarios. The results show that observers can indeed determine the relative location of a person in a scene with a significantly higher accuracy when they perform with fused images, compared to the individual image modalities. The MIT colour fusion scheme yields the best overall performance (i.e. an accuracy that is significantly higher than that obtained with images fused according to the TNO scheme and with the original images). The TNO colour fused images result in an accuracy that is in between the accuracy obtained for the individual (thermal and visual) images and the accuracy obtained with images fused according to the MIT scheme. The TNO scheme did not involve the any image enhancement on the original images. The inclusion of a standard contrast stretching operation prior to the fusion process may lead to an improved observer perrformance. The present results prove that even a very simple colour fusion scheme yield an improved observer performance in a relative spatial localization task.

The difference between the localisation performance with the colour fused images from both schemes and with their luminance components (the derived graylevel fused images) is not significant. However, for both schemes the detection performance is significantly better (less misses) for the colour fused images than for the graylevel fused images. This proves that colour fused images are easier to visually segment than graylevel fused images. Therefore, different tasks like navigation and orienteering, that probably depend on a correct scene segmentation to obtain a sufficient amount of situational awareness, may benefit from a colour fused image representation. Further research, preferably involving dynamic scenarios, is needed to test the hypothesis that colour image fusion schemes can boost observer performance in these tasks.

The present results indicate that the fusion of thermal and visual imagery can lead to an improved observer performance for localization tasks. It is therefore likely that the fusion of thermal and low-light level imagery may yield an even better observer performance over an even larger time frame.

5 CONCLUSIONS

The fusion of thermal and visual images registered around sunrise results in composite images with an increased amount of detail that clearly represent all details in their correct spatial context.

Colour fused images yield a significantly better detection performance than graylevel fused images.

Observers can indeed determine the relative location of a person in a scene with a significantly higher accuracy when they perform with fused images, compared to the individual image modalities.

The MIT colour fusion scheme yields the best overall performance (i.e. an accuracy that is significantly higher than that obtained with images fused according to the TNO scheme and with the original images).

The present results prove that even a very simple colour fusion method like the TNO scheme yields an improved observer performance in a relative spatial localization task.

Acknowledgement

We thank Klamer Schutte and Robin de Rooy of the TNO Physics and Electronics Laboratory for their help with the image registration.
We are grateful to Prof. Dr. Allen Waxman and Mario Aguilar, Machine Intelligence Group, MIT Lincoln Laboratory, Lexington, USA, for providing us with color and graylevel fused images.

REFERENCES

Buimer, A.R. (1993). Scenario's for multi-spectral image fusion (in Dutch). Memo 21 December 1993, Sectie Plannen, Opleidings Centrum Infanterie (OCI), Harderwijk, The Netherlands.

Carts-Powell, Y. (1996). Fusing CCD and IR images creates color night vision. *Laser Focus World, 32(5),* 32–36.

DeValois, K.K., Lakshminarayanan, V., Nygaard, R., Schlussel, S., & Sladky, J. (1990). Discrimination of relative spatial position. *Vision Research, 30,* 1649–1660.

Fink, W. (1976). Image coloration as an interpretation aid. *Proceedings SPIE/OSA Image Processing, 74,* 209–215.

Gouras, P. (1991). Color vision. In E.R. Kandel, J.H. Schwartz & T.M. Jessell (Eds.), *Principles of Neural Science, 3rd ed.* (pp. 467–480). Oxford: Elsevier Science Publishers.

Gove, A.N., Cunningham, R.K. & Waxman, A.M. (1996). Opponent-color visual processing applied to multispectral infrared imagery. *Proceedings of the 1996 Meeting of the IRIS Specialty Group on Passive Sensors* (in press). Monterey, CA.

Grossberg, S. (1988). *Neural networks and natural intelligence.* MIT Press, Cambridge, MA, USA.

NATO (1993). *Proceedings of the AC/243 (Panel 3) and (Panel 4) Symposium on Multi-sensors and Sensor Data Fusion, 8–12 November 1993,* Brussels.

NATO (1994). Joint Symposium on Multisensors and Sensor Data Fusion. NATO Technical Proceedings AC/243-TP/8, 5-1–517.

Newman, E.A. & Hartline, P.H. (1981). Integration of visual and infrared information in bimodal neurons of the rattlesnake optic tectum. *Science, 213,* 789–791.

Newman, E.A. & Hartline, P.H. (1982). The infrared "vision" of snakes. *Scientific American, 246,* 116–127.

Pratt, W.K. (1991). *Digital Image Processing.* New York, USA: Wiley.

Savoye, E.D, Waxman, A.M., Buss, J., Hawkins, H. & Campana, S.B. (1996). Charge coupled devices and visible/IR fusion for night vision. *Proceedings of Night Vision '96.* London, UK.

Schiller, P. (1992). The ON and OFF channels of the visual system. *Trends in Neuroscience, 15,* 86–92.

Schiller, P. & Logothesis, N.K. (1990). The color-opponent and broad-band channels of the primate visual system. *Trends in Neuroscience, 13,* 392–398.

Schwering, P. (1995). *Verslag van de 36e vergadering van AC/243 (Panel 3/RSG.9) 'On Image Processing', 2–6 oktober 1995, Den Haag, Nederland* (TNO Report BV-1995-270). The Hague, The Netherlands: TNO Physics and Electronics Lab.

Sévigny, L. (1996). *Multisensor image fusion for detection and recognition of targets in the battlefield of the future* (Progress Report Canada, NATO AC/243, Panel 3, RSG.9 37th meeting). Quèbec, Canada: Defense Research Establishment Valcartier.

Toet, A. & Walraven, J. (1996). New false colour mapping for image fusion. *Optical Engineering, 35(3),* 650–658.

Uttal, W.R., Baruch, T. & Allen, L. (1995). Dichoptic and physical information combination: a comparison. *Perception, 24,* 351–362.

Waxman, A.M., Carrick, J.E., Fay, D.A., Racamato, J.P., Augilar, M., and Savoye, E.D. (1996). Electronic imaging aids for night driving: low-light CCD, thermal IR, and color fused visible/IR. *Proceedings of the SPIE Conference on Transportation Sensors and Controls, SPIE-2902.*

Waxman, A.M., Fay, D.A., Gove, A.N., Seibert, M., Racamoto, J.P., Carrick, J.E. & Savoye, E.D. (1995). Color night vision: fusion of intensified visible and thermal IR imagery. *Proceedings of the SPIE Conference on 2463 on Synthetic Vision for Vehicle Guidance and Control, vol. SPIE-2463*, 58–68.

Waxman, A.M., Gove, A.N. & Cunningham, R.K. (1996). Opponent-Color Visual Processing Applied to Multispectral Infrared Imagery. *Proceedings of 1996 Meeting of the IRIS Specialty Group on Passive Sensors, II*, 247–262. Ann Arbor, US: Infrared Information Analysis Center, ERIM.

Waxman, A.M., Gove, A.N., Fay, D.A., Racamoto, J.P., Carrick, J.E., Seibert, M. & Savoye, E.D. (1996a). Color night vision: opponent processing in the fusion of visible and IR imagery. *Neural Networks, 9(6)*, in press.

Waxman, A.M., Gove, A.N., Fay, D.A., Racamoto, J.P., Carrick, J.E., Seibert, M., Savoye, E.D., Burke, B.E., Reich, R.K., McGonagle, W.H., and Craig, D.M. (1996). Solid state color night vision: fusion of low-light visible and thermal IR imagery. *Proceedings of the 1996 Meeting of the IRIS Specialty Group on Passive Sensors, II*, 263–280. Ann Arbor: Infrared Information Analysis Center, ERIM.

Waxman, A.M., Gove, A.N., Seibert, M., Fay, D.A., Carrick, J.E., Racamoto, J.P., Savoye, E.D., Burke, B.E., Reich, R.K., McGonagle, W.H. & Craig, D.M. (1996). Progress on color night vision: visible/IR fusion, perception & search, and low-light CCD imaging. *Proceedings of the SPIE Conference on Enhanced and Synthetic Vision, vol. SPIE-2736*, 96–107.

Soesterberg, 31 October, 1996

Dr. A. Toet
(1st author, projectleader)

VERZENDLIJST

1. Directeur M&P DO

2. Directie Wetenschappelijk Onderzoek en Ontwikkeling Defensie

3. { Hoofd Wetenschappelijk Onderzoek KL
Plv. Hoofd Wetenschappelijk Onderzoek KL

4. Hoofd Wetenschappelijk Onderzoek KLu

5. { Hoofd Wetenschappelijk Onderzoek KM
Plv. Hoofd Wetenschappelijk Onderzoek KM

6, 7 en 8. Bibliotheek KMA, Breda

9. LKol. G.H. Bakema, Hoofd Bureau Sectie Externe Plannen Infanterie,
KCEN OCMAN, Amersfoort

10 t/m 18. Leden WOOST

Extra exemplaren van dit rapport kunnen worden aangevraagd door tussenkomst van de HWOs of de DWOO.